

**Compatible multi-channel coding/decoding**Field of the invention

5

The present invention relates to an apparatus and a method for processing a multi-channel audio signal and, in particular, to an apparatus and a method for processing a multi-channel audio signal in a stereo-compatible manner.

10

Background of the Invention and Prior Art

In recent times, the multi-channel audio reproduction technique is becoming more and more important. This may be due to the fact that audio compression/encoding techniques such as the well-known mp3 technique have made it possible to distribute audio records via the Internet or other transmission channels having a limited bandwidth. The mp3 coding technique has become so famous because of the fact that it allows distribution of all the records in a stereo format, i.e., a digital representation of the audio record including a first or left stereo channel and a second or right stereo channel.

25

Nevertheless, there are basic shortcomings of conventional two-channel sound systems. Therefore, the surround technique has been developed. A recommended multi-channel-surround representation includes, in addition to the two stereo channels L and R, an additional center channel C and two surround channels Ls, Rs. This reference sound format is also referred to as three/two-stereo, which means three front channels and two surround channels. Generally, five

transmission channels are required. In a playback environment, at least five speakers at the respective five different places are needed to get an optimum sweet spot in a certain distance from the five well-placed loudspeakers.

5

Several techniques are known in the art for reducing the amount of data required for transmission of a multi-channel audio signal. Such techniques are called joint stereo techniques. To this end, reference is made to Fig. 10, which 10 shows a joint stereo device 60. This device can be a device implementing e.g. intensity stereo (IS) or binaural cue coding (BCC). Such a device generally receives - as an input - at least two channels (CH<sub>1</sub>, CH<sub>2</sub>, ... CH<sub>n</sub>), and outputs a single carrier channel and parametric data. The parametric data are defined such that, in a decoder, an approximation 15 of an original channel (CH<sub>1</sub>, CH<sub>2</sub>, ... CH<sub>n</sub>) can be calculated.

Normally, the carrier channel will include subband samples, 20 spectral coefficients, time domain samples etc, which provide a comparatively fine representation of the underlying signal, while the parametric data do not include such samples of spectral coefficients but include control parameters for controlling a certain reconstruction algorithm 25 such as weighting by multiplication, time shifting, frequency shifting, ... The parametric data, therefore, include only a comparatively coarse representation of the signal or the associated channel. Stated in numbers, the amount of data required by a carrier channel will be in the range of 30 60 - 70 kbit/s, while the amount of data required by parametric side information for one channel will be in the range of 1,5 - 2,5 kbit/s. An example for parametric data

are the well-known scale factors, intensity stereo information or binaural cue parameters as will be described below.

Intensity stereo coding is described in AES preprint 3799,  
5 "Intensity Stereo Coding", J. Herre, K. H. Brandenburg, D.  
Lederer, February 1994, Amsterdam. Generally, the concept  
of intensity stereo is based on a main axis transform to be  
applied to the data of both stereophonic audio channels. If  
most of the data points are concentrated around the first  
10 principle axis, a coding gain can be achieved by rotating  
both signals by a certain angle prior to coding. This is,  
however, not always true for real stereophonic production  
techniques. Therefore, this technique is modified by ex-  
cluding the second orthogonal component from transmission  
15 in the bit stream. Thus, the reconstructed signals for the  
left and right channels consist of differently weighted or  
scaled versions of the same transmitted signal. Neverthe-  
less, the reconstructed signals differ in their amplitude  
but are identical regarding their phase information. The  
20 energy-time envelopes of both original audio channels, how-  
ever, are preserved by means of the selective scaling op-  
eration, which typically operates in a frequency selective  
manner. This conforms to the human perception of sound at  
high frequencies, where the dominant spatial cues are de-  
25 termined by the energy envelopes.

Additionally, in practically implementations, the transmit-  
ted signal, i.e. the carrier channel is generated from the  
30 sum signal of the left channel and the right channel in-  
stead of rotating both components. Furthermore, this proc-  
essing, i.e., generating intensity stereo parameters for  
performing the scaling operation, is performed frequency  
selective, i.e., independently for each scale factor band,

i.e., encoder frequency partition. Preferably, both channels are combined to form a combined or "carrier" channel, and, in addition to the combined channel, the intensity stereo information is determined which depend on the energy of the first channel, the energy of the second channel or the energy of the combined or channel.

The BCC technique is described in AES convention paper 5574, "Binaural cue coding applied to stereo and multi-channel audio compression", C. Faller, F. Baumgarte, May 10 2002, Munich. In BCC encoding, a number of audio input channels are converted to a spectral representation using a DFT based transform with overlapping windows. The resulting uniform spectrum is divided into non-overlapping partitions 15 each having an index. Each partition has a bandwidth proportional to the equivalent rectangular bandwidth (ERB). The inter-channel level differences (ICLD) and the inter-channel time differences (ICTD) are estimated for each partition for each frame k. The ICLD and ICTD are quantized 20 and coded resulting in a BCC bit stream. The inter-channel level differences and inter-channel time differences are given for each channel relative to a reference channel. Then, the parameters are calculated in accordance with prescribed formulae, which depend on the certain partitions of 25 the signal to be processed.

At a decoder-side, the decoder receives a mono signal and the BCC bit stream. The mono signal is transformed into the frequency domain and input into a spatial synthesis block, 30 which also receives decoded ICLD and ICTD values. In the spatial synthesis block, the BCC parameters (ICLD and ICTD) values are used to perform a weighting operation of the mono signal in order to synthesize the multi-channel sig-

nals, which, after a frequency/time conversion, represent a reconstruction of the original multi-channel audio signal.

- In case of BCC, the joint stereo module 60 is operative to  
5 output the channel side information such that the parametric channel data are quantized and encoded ICLD or ICTD parameters, wherein one of the original channels is used as the reference channel for coding the channel side information.

- 10 Normally, the carrier channel is formed of the sum of the participating original channels.

- Naturally, the above techniques only provide a mono representation for a decoder, which can only process the carrier channel, but is not able to process the parametric data for generating one or more approximations of more than one input channel.

- 20 To transmit the five channels in a compatible way, i.e., in a bitstream format, which is also understandable for a normal stereo decoder, the so-called matrixing technique has been used as described in "MUSICAM surround: a universal multi-channel coding system compatible with ISO 11172-3",  
25 G. Theile and G. Stoll, AES preprint 3403, October 1992, San Francisco. The five input channels L, R, C, Ls, and Rs are fed into a matrixing device performing a matrixing operation to calculate the basic or compatible stereo channels Lo, Ro, from the five input channels. In particular,  
30 these basic stereo channels Lo/Ro are calculated as set out below:

$$Lo = L + xC + yLs$$

$$Ro = R + xC + yRs$$

- x and y are constants. The other three channels C, Ls, Rs  
5 are transmitted as they are in an extension layer, in addition  
to a basic stereo layer, which includes an encoded  
version of the basic stereo signals Lo/Ro. With respect to  
the bitstream, this Lo/Ro basic stereo layer includes a  
header, information such as scale factors and subband sam-  
10 ples. The multi-channel extension layer, i.e., the central  
channel and the two surround channels are included in the  
multi-channel extension field, which is also called ancil-  
lary data field.
- 15 At a decoder-side, an inverse matrixing operation is per-  
formed in order to form reconstructions of the left and  
right channels in the five-channel representation using the  
basic stereo channels Lo, Ro and the three additional chan-  
nels. Additionally, the three additional channels are de-  
20 coded from the ancillary information in order to obtain a  
decoded five-channel or surround representation of the  
original multi-channel audio signal.

Another approach for multi-channel encoding is described in  
25 the publication "Improved MPEG-2 audio multi-channel encod-  
ing", B. Grill, J. Herre, K. H. Brandenburg, E. Eberlein,  
J. Koller, J. Mueller, AES preprint 3865, February 1994,  
Amsterdam, in which, in order to obtain backward compati-  
bility, backward compatible modes are considered. To this  
30 end, a compatibility matrix is used to obtain two so-called  
downmix channels Lc, Rc from the original five input chan-  
nels. Furthermore, it is possible to dynamically select  
the three auxiliary channels transmitted as ancillary data.

In order to exploit stereo irrelevancy, a joint stereo technique is applied to groups of channels, e. g. the three front channels, i.e., for the left channel, the right channel and the center channel. To this end, these three channels are combined to obtain a combined channel. This combined channel is quantized and packed into the bitstream. Then, this combined channel together with the corresponding joint stereo information is input into a joint stereo decoding module to obtain joint stereo decoded channels, i.e., a joint stereo decoded left channel, a joint stereo decoded right channel and a joint stereo decoded center channel. These joint stereo decoded channels are, together with the left surround channel and the right surround channel input into a compatibility matrix block to form the first and the second downmix channels  $L_c$ ,  $R_c$ . Then, quantized versions of both downmix channels and a quantized version of the combined channel are packed into the bitstream together with joint stereo coding parameters.

Using intensity stereo coding, therefore, a group of independent original channel signals is transmitted within a single portion of "carrier" data. The decoder then reconstructs the involved signals as identical data, which are rescaled according to their original energy-time envelopes. Consequently, a linear combination of the transmitted channels will lead to results, which are quite different from the original downmix. This applies to any kind of joint stereo coding based on the intensity stereo concept. For a coding system providing compatible downmix channels, there is a direct consequence: The reconstruction by dematrixing, as described in the previous publication, suffers from artifacts caused by the imperfect reconstruction. Using a so-

- called joint stereo predistortion scheme, in which a joint stereo coding of the left, the right and the center channels is performed before matrixing in the encoder, alleviates this problem. In this way, the dematrixing scheme for reconstruction introduces fewer artifacts, since, on the encoder-side, the joint stereo decoded signals have been used for generating the downmix channels. Thus, the imperfect reconstruction process is shifted into the compatible downmix channels Lc and Rc, where it is much more likely to be masked by the audio signal itself.
- Although such a system has resulted in fewer artifacts because of dematrixing on the decoder-side, it nevertheless has some drawbacks. A drawback is that the stereo-compatible downmix channels Lc and Rc are derived not from the original channels but from intensity stereo coded/decoded versions of the original channels. Therefore, data losses because of the intensity stereo coding system are included in the compatible downmix channels. A stereo-only decoder, which only decodes the compatible channels rather than the enhancement intensity stereo encoded channels, therefore, provides an output signal, which is affected by intensity stereo induced data losses.
- 25 Additionally, a full additional channel has to be transmitted besides the two downmix channels. This channel is the combined channel, which is formed by means of joint stereo coding of the left channel, the right channel and the center channel. Additionally, the intensity stereo information to reconstruct the original channels L, R, C from the combined channel also has to be transmitted to the decoder. At the decoder, an inverse matrixing, i.e., a dematrixing operation is performed to derive the surround channels from

- the two downmix channels. Additionally, the original left, right and center channels are approximated by joint stereo decoding using the transmitted combined channel and the transmitted joint stereo parameters. It is to be noted that  
5 the original left, right and center channels are derived by joint stereo decoding of the combined channel.

Summary of the Invention

10

It is the object of the present invention to provide a concept for a bit-efficient and artifact-reduced processing or inverse processing of a multi-channel audio signal.

- 15 In accordance with a first aspect of the present invention, this object is achieved by an apparatus for processing a multi-channel audio signal, the multi-channel audio signal having at least three original channels, comprising: means for providing a first downmix channel and a second downmix  
20 channel, the first and the second downmix channels being derived from the original channels; means for calculating channel side information for a selected original channel of the original signals, the means for calculating being operative to calculate the channel side information such that  
25 a downmix channel or a combined downmix channel including the first and the second downmix channel, when weighted using the channel side information, results in an approximation of the selected original channel; and means for generating output data, the output data including the channel side information, the first downmix channel or a signal derived from the first downmix channel and the second downmix  
30 channel or a signal derived from the second downmix channel.

- In accordance with a second aspect of the present invention, this object is achieved by a method of processing a multi-channel audio signal, the multi-channel audio signal
- 5 having at least three original channels, comprising: providing a first downmix channel and a second downmix channel, the first and the second downmix channels being derived from the original channels; calculating channel side information for a selected original channel of the original
- 10 signals such that a downmix channel or a combined downmix channel including the first and the second downmix channel, when weighted using the channel side information, results in an approximation of the selected original channel; and generating output data, the output data including the channel side information, the first downmix channel or a signal derived from the first downmix channel and the second downmix channel or a signal derived from the second downmix channel.
- 15
- 20 In accordance with a third aspect of the present invention, this object is achieved by an apparatus for inverse processing of input data, the input data including channel side information, a first downmix channel or a signal derived from the first downmix channel and a second downmix channel or a signal derived from the second downmix channel,
- 25 wherein the first downmix channel and the second downmix channel are derived from at least three original channels of a multi-channel audio signal, and wherein the channel side information are calculated such that a downmix channel
- 30 or a combined downmix channel including the first downmix channel and the second downmix channel, when weighted using the channel side information, results in an approximation of the selected original channel, the apparatus comprising:

- an input data reader for reading the input data to obtain the first downmix channel or a signal derived from the first downmix channel and the second downmix channel or a signal derived from the second downmix channel and the channel side information; and a channel reconstructor for reconstructing the approximation of the selected original channel using the channel side information and the downmix channel or the combined downmix channel to obtain the approximation of the selected original channel.
- In accordance with a fourth aspect of the present invention, this object is achieved by a method of inverse processing of input data, the input data including channel side information, a first downmix channel or a signal derived from the first downmix channel and a second downmix channel, or a signal derived from the second downmix channel, wherein the first downmix channel and the second downmix channel are derived from at least three original channels of a multi-channel audio signal, and wherein the channel side information are calculated such that a downmix channel or a combined downmix channel including the first downmix channel and the second downmix channel, when weighted using the channel side information, results in an approximation of the selected original channel, the method comprising:
- reading the input data to obtain the first downmix channel or a signal derived from the first downmix channel and the second downmix channel or a signal derived from the second downmix channel and the channel side information; and reconstructing the approximation of the selected original channel using the channel side information and the downmix channel or the combined downmix channel to obtain the approximation of the selected original channel.

In accordance with a fifth aspect and a sixth aspect of the present invention, this object is achieved by a computer program including the method of processing or the method of inverse processing.

5       The present invention is based on the finding that an efficient and artifact-reduced encoding of multi-channel audio signal is obtained, when two downmix channels preferably representing the left and right stereo channels, are packed  
10      into output data.

Inventively, parametric channel side information for one or more of the original channels are derived such that they relate to one of the downmix channels rather than, as in the prior art, to an additional "combined" joint stereo channel. This means that the parametric channel side information are calculated such that, on a decoder side, a channel reconstructor uses the channel side information and one of the downmix channels or a combination of the downmix 20      channels to reconstruct an approximation of the original audio channel, to which the channel side information is assigned.

The inventive concept is advantageous in that it provides a 25      bit-efficient multi-channel extension such that a multi-channel audio signal can be played at a decoder.

Additionally, the inventive concept is backward compatible, since a lower scale decoder, which is only adapted for two-channel processing, can simply ignore the extension information, i.e., the channel side information. The lower scale 30      decoder can only play the two downmix channels to obtain a stereo representation of the original multi-channel audio

signal. A higher scale decoder, however, which is enabled for multi-channel operation, can use the transmitted channel side information to reconstruct approximations of the original channels.

- 5       The present invention is advantageous in that it is bit-efficient, since, in contrast to the prior art, no additional carrier channel beyond the first and second downmix channels  $L_c, R_c$  is required. Instead, the channel side information are related to one or both downmix channels. This means that the downmix channels themselves serve as a carrier channel, to which the channel side information are combined to reconstruct an original audio channel. This means that the channel side information are preferably parametric side information, i.e., information which do not include any subband samples or spectral coefficients. Instead, the parametric side information are information used for weighting (in time and/or frequency) the respective downmix channel or the combination of the respective downmix channels to obtain a reconstructed version of a selected original channel.
- 10      15      20

In a preferred embodiment of the present invention, a backward compatible coding of a multi-channel signal based on a compatible stereo signal is obtained. Preferably, the compatible stereo signal (downmix signal) is generated using matrixing of the original channels of multi-channel audio signal.

- 25      30      Inventively, channel side information for a selected original channel is obtained based on joint stereo techniques such as intensity stereo coding or binaural cue coding. Thus, at the decoder side, no dematrixing operation has to

- be performed. The problems associated with dematrixing, i.e., certain artifacts related to an undesired distribution of quantization noise in dematrixing operations, are avoided. This is due to the fact that the decoder uses a channel reconstructor, which reconstructs an original signal, by using one of the downmix channels or a combination of the downmix channels and the transmitted channel side information.
- 5
- 10 Preferably, the inventive concept is applied to a multi-channel audio signal having five channels. These five channels are a left channel L, a right channel R, a center channel C, a left surround channel Ls, and a right surround channel Rs. Preferably, downmix channels are stereo compatible downmix channels Ls and Rs, which provide a stereo representation of the original multi-channel audio signal.
- 15
- In accordance with the preferred embodiment of the present invention, for each original channel, channel side information are calculated at an encoder side packed into output data. Channel side information for the original left channel are derived using the left downmix channel. Channel side information for the original left surround channel are derived using the left downmix channel. Channel side information for the original right channel are derived from the right downmix channel. Channel side information for the original right surround channel are derived from the right downmix channel.
- 20
- 25
- 30 In accordance with the preferred embodiment of the present invention, channel information for the original center channel are derived using the first downmix channel as well as the second downmix channel, i.e., using a combination of

the two downmix channels. Preferably, this combination is a summation.

Thus, the groupings, i.e., the relation between the channel side information and the carrier signal, i.e., the used downmix channel for providing channel side information for a selected original channel are such that, for optimum quality, a certain downmix channel is selected, which contains the highest possible relative amount of the respective original multi-channel signal which is represented by means of channel side information. As such a joint stereo carrier signal, the first and the second downmix channels are used. Preferably, also the sum of the first and the second downmix channels can be used. Naturally, the sum of the first and second downmix channels can be used for calculating channel side information for each of the original channels. Preferably, however, the sum of the downmix channels is used for calculating the channel side information of the original center channel in a surround environment, such as five channel surround, seven channel surround, 5.1 surround or 7.1 surround. Using the sum of the first and second downmix channels is especially advantageous, since no additional transmission overhead has to be performed. This is due to the fact that both downmix channels are presented at the decoder such that summing of these downmix channels can easily be performed at the decoder without requiring any additional transmission bits.

Preferably, the channel side information forming the multi-channel extension are input into the output data bit stream in a compatible way such that a lower scale decoder simply ignores the multi-channel extension data and only provides a stereo representation of the multi-channel audio signal.

Nevertheless, a higher scale encoder not only uses two downmix channels, but, in addition, employs the channel side information to reconstruct a full multi-channel representation of the original audio signal.

5 An inventive decoder is operative to firstly decode both downmix channels and to read the channel side information for the selected original channels. Then, the channel side information and the downmix channels are used to reconstruct approximations of the original channels. To this end, preferably no dematrixing operation at all is performed. This means that, in this embodiment, each of the e.g. five original input channels are reconstructed using e.g. five sets of different channel side information. In the 10 decoder, the same grouping as in the encoder is performed for calculating the reconstructed channel approximation. In a five-channel surround environment, this means that, for reconstructing the original left channel, the left downmix channel and the channel side information for the left channel are used. To reconstruct the original right channel, the right downmix channel and the channel side information for the right channel are used. To reconstruct the original left surround channel, the left downmix channel and the channel side information for the left surround channel are 15 used. To reconstruct the original right surround channel, the channel side information for the right surround channel and the right downmix channel are used. To reconstruct the original center channel, a combined channel formed from the first downmix channel and the second downmix channel and the center channel side information are used.

20

25

30

Naturally, it is also possible, to replay the first and second downmix channels as the left and right channels such

that only three sets (out of e. g. five) of channel side information parameters have to be transmitted. This is, however, only advisable in situations, where there are less stringent rules with respect to quality. This is due to the fact that, normally, the left downmix channel and the right downmix channel are different from the original left channel or the original right channel. Only in situations, where one can not afford to transmit channel side information for each of the original channels, such processing is advantageous.

Brief Description of the Drawings

Preferred embodiments of the present invention are subsequently discussed with reference to the attached figures, in which:

Fig. 1 is a block diagram of a preferred embodiment of the inventive encoder;

Fig. 2 is a block diagram of a preferred embodiment of the inventive decoder;

Fig. 3A is a block diagram for a preferred implementation of the means for calculating to obtain frequency selective channel side information;

Fig. 3B is a preferred embodiment of a calculator implementing joint stereo processing such as intensity coding or binaural cue coding;

Fig. 4 illustrates another preferred embodiment of the means for calculating channel side information, in which the channel side information are gain factors;

5 Fig. 5 illustrates a preferred embodiment of an implementation of the decoder, when the encoder is implemented as in Fig. 4;

10 Fig. 6 illustrates a preferred implementation of the means for providing the downmix channels;

15 Fig. 7 illustrates groupings of original and downmix channels for calculating the channel side information for the respective original channels;

Fig. 8 illustrates another preferred embodiment of an inventive encoder;

20 Fig. 9 illustrates another implementation of an inventive decoder; and

Fig. 10 illustrates a prior art joint stereo encoder.

25

Detailed Description of Preferred Embodiments

Fig. 1 shows an apparatus for processing a multi-channel audio signal 10 having at least three original channels such as R, L and C. Preferably, the original audio signal has more than three channels, such as five channels in the surround environment, which is illustrated in Fig. 1. The five channels are the left channel L, the right channel R,

the center channel C, the left surround channel Ls and the right surround channel Rs. The inventive apparatus includes means 12 for providing a first downmix channel Lc and a second downmix channel Rc, the first and the second downmix channels being derived from the original channels. For deriving the downmix channels from the original channels, there exist several possibilities. One possibility is to derive the downmix channels Lc and Rc by means of matrixing the original channels using a matrixing operation as illustrated in Fig. 6. This matrixing operation is performed in the time domain.

The matrixing parameters a, b and t are selected such that they are lower than or equal to 1. Preferably, a and b are 15 0.7 or 0.5. The overall weighting parameter t is preferably chosen such that channel clipping is avoided. Alternatively, as it is indicated in Fig. 1, the downmix channels Lc and Rc can also be externally supplied. This may be done, when the downmix channels Lc and Rc are the 20 result of a "hand mixing" operation. In this scenario, a sound engineer mixes the downmix channels by himself rather than by using an automated matrixing operation. The sound engineer performs creative mixing to get optimized downmix channels Lc and Rc which give the best possible stereo representation of the original multi-channel audio signal.

In case of an external supply of the downmix channels, the means for providing does not perform a matrixing operation but simply forwards the externally supplied downmix channels to a subsequent calculating means 14.

The calculating means 14 is operative to calculate the channel side information such as  $l_i$ ,  $l_{si}$ ,  $r_i$  or  $r_{si}$  for se-

selected original channels such as L, Ls, R or Rs, respectively. In particular, the means 14 for calculating is operative to calculate the channel side information such that a downmix channel, when weighted using the channel side information, results in an approximation of the selected original channel.

- Alternatively or additionally, the means for calculating channel side information is further operative to calculate 10 the channel side information for a selected original channel such that a combined downmix channel including a combination of the first and second downmix channels, when weighted using the calculated channel side information results in an approximation of the selected original channel.
- 15 To show this feature in the figure, an adder 14a and a combined channel side information calculator 14b are shown.

It is clear for those skilled in the art that these elements do not have to be implemented as distinct elements. 20 Instead, the whole functionality of the blocks 14, 14a, and 14b can be implemented by means of a certain processor which may be a general purpose processor or any other means for performing the required functionality.

- 25 Additionally, it is to be noted here that channel signals being subband samples or frequency domain values are indicated in capital letters. Channel side information are, in contrast to the channels themselves, indicated by small letters. The channel side information  $c_i$  is, therefore, the 30 channel side information for the original center channel C.

The channel side information as well as the downmix channels Lc and Rc or an encoded version Lc' and Rc' as pro-

duced by an audio encoder 16 are input into an output data formatter 18. Generally, the output data formatter 18 acts as means for generating output data, the output data including the channel side information for at least one original channel, the first downmix channel or a signal derived from the first downmix channel (such as an encoded version thereof) and the second downmix channel or a signal derived from the second downmix channel (such as an encoded version thereof).

The output data or output bitstream 20 can then be transmitted to a bitstream decoder or can be stored or distributed. Preferably, the output bitstream 20 is a compatible bitstream which can also be read by a lower scale decoder not having a multi-channel extension capability. Such lower scale encoders such as most existing normal state of the art mp3 decoders will simply ignore the multi-channel extension data, i.e., the channel side information. They will only decode the first and second downmix channels to produce a stereo output. Higher scale decoders, such as multi-channel enabled decoders will read the channel side information and will then generate an approximation of the original audio channels such that a multi-channel audio impression is obtained.

Fig. 8 shows a preferred embodiment of the present invention in the environment of five channel surround / mp3. Here, it is preferred to write the surround enhancement data into the ancillary data field in the standardized mp3 bit stream syntax such that an "mp3 surround" bit stream is obtained.

Fig. 2 shows an illustration of an inventive decoder acting as an apparatus for inverse processing input data received at an input data port 22. The data received at the input data port 22 is the same data as output at the output data port 20 in Fig. 1. Alternatively, when the data are not transmitted via a wired channel but via a wireless channel, the data received at data input port 22 are data derived from the original data produced by the encoder.

- 10 The decoder input data are input into a data stream reader 24 for reading the input data to finally obtain the channel side information 26 and the left downmix channel 28 and the right downmix channel 30. In case the input data includes encoded versions of the downmix channels, which corresponds 15 to the case, in which the audio encoder 16 in Fig. 1 is present, the data stream reader 24 also includes an audio decoder, which is adapted to the audio encoder used for decoding the downmix channels. In this case, the audio decoder, which is part of the data stream reader 24, is operative to generate the first downmix channel  $L_c$  and the second downmix channel  $R_c$ , or, stated more exactly, a decoded version of those channels. For ease of description, a distinction between signals and decoded versions thereof is 20 made where explicitly stated.
- 25 The channel side information 26 and the left and right downmix channels 28 and 30 output by the data stream reader 24 are fed into a multi-channel reconstructor 32 for providing a reconstructed version 34 of the original audio signals, which can be played by means of a multi-channel player 36. In case the multi-channel reconstructor is operative in the frequency domain, the multi-channel player 30 36 will receive frequency domain input data, which have to

be in a certain way decoded such as converted into the time domain before playing them. To this end, the multi-channel player 36 may also include decoding facilities.

- 5 It is to be noted here that a lower scale decoder will only have the data stream reader 24, which only outputs the left and right downmix channels 28 and 30 to a stereo output 38. An enhanced inventive decoder will, however, extract the channel side information 26 and use these side information 10 and the downmix channels 28 and 30 for reconstructing reconstructed versions 34 of the original channels using the multi-channel reconstructor 32.

Fig. 3A shows an embodiment of the inventive calculator 14 for calculating the channel side information, which an audio encoder on the one hand and the channel side information calculator on the other hand operate on the same spectral representation of multi-channel signal. Fig. 1, however, shows the other alternative, in which the audio encoder on the one hand and the channel side information calculator on the other hand operate on different spectral representations of the multi-channel signal. When computing resources are not as important as audio quality, the Fig. 1 alternative is preferred, since filterbanks individually 20 optimized for audio encoding and side information calculation can be used. When, however, computing resources are an issue, the Fig. 3A alternative is preferred, since this alternative requires less computing power because of a shared utilization of elements.

- 30 The device shown in Fig. 3A is operative for receiving two channels A, B. The device shown in Fig. 3A is operative to calculate a side information for channel B such that using

this channel side information for the selected original channel B, a reconstructed version of channel B can be calculated from the channel signal A. Additionally, the device shown in Fig. 3A is operative to form frequency domain 5 channel side information, such as parameters for weighting (by multiplying or time processing as in BCC coding e. g.) spectral values or subband samples. To this end, the inventive calculator includes windowing and time/frequency conversion means 140a to obtain a frequency representation of 10 channel A at an output 140b or a frequency domain representation of channel B at an output 140c.

In the preferred embodiment, the side information determination (by means of the side information determination 15 means 140f) is performed using quantized spectral values. Then, a quantizer 140d is also present which preferably is controlled using a psychoacoustic model having a psycho-acoustic model control input 140e. Nevertheless, a quantizer is not required, when the side information determination 20 means 140c uses a non-quantized representation of the channel A for determining the channel side information for channel B.

In case the channel side information for channel B are calculated by means of a frequency domain representation of 25 the channel A and the frequency domain representation of the channel B, the windowing and time/frequency conversion means 140a can be the same as used in a filterbank-based audio encoder. In this case, when AAC (ISO/IEC 13818-3) is considered, means 140a is implemented as an MDCT filter 30 bank (MDCT = modified discrete cosine transform) with 50% overlap-and-add functionality.

In such a case, the quantizer 140d is an iterative quantizer such as used when mp3 or AAC encoded audio signals are generated. The frequency domain representation of channel A, which is preferably already quantized can then be directly used for entropy encoding using an entropy encoder 140g, which may be a Huffman based encoder or an entropy encoder implementing arithmetic encoding.

When compared to Fig. 1, the output of the device in Fig. 10 3A is the side information such as  $l_1$  for one original channel (corresponding to the side information for B at the output of device 140f). The entropy encoded bitstream for channel A corresponds to e. g. the encoded left downmix channel  $Lc'$  at the output of block 16 in Fig. 1. From Fig. 15 3A it becomes clear that element 14 (Fig. 1), i.e., the calculator for calculating the channel side information and the audio encoder 16 (Fig. 1) can be implemented as separate means or can be implemented as a shared version such that both devices share several elements such as the MDCT 20 filter bank 140a, the quantizer 140e and the entropy encoder 140g. Naturally, in case one needs a different transform etc. for determining the channel side information, then the encoder 16 and the calculator 14 (Fig. 1) will be implemented in different devices such that both elements do 25 not share the filter bank etc.

Generally, the actual determinator for calculating the side information (or generally stated the calculator 14) may be implemented as a joint stereo module as shown in Fig.3B, 30 which operates in accordance with any of the joint stereo techniques such as intensity stereo coding or binaural cue coding.

In contrast to such prior art intensity stereo encoders, the inventive determination means 140f does not have to calculate the combined channel. The "combined channel" or carrier channel, as one can say, already exists and is the 5 left compatible downmix channel Lc or the right compatible downmix channel Rc or a combined version of these downmix channels such as Lc + Rc. Therefore, the inventive device 10 only has to calculate the scaling information for scaling the respective downmix channel such that the energy/time envelope of the respective selected original channel is obtained, when the downmix channel is weighted 15 using the scaling information or, as one can say, the intensity directional information.

15 Therefore, the joint stereo module 140f in Fig 3B is illustrated such that it receives, as an input, the "combined" channel A, which is the first or second downmix channel or a combination of the downmix channels, and the original selected channel. This module, naturally, outputs the "combined" channel A and the joint stereo parameters as channel 20 side information such that, using the combined channel A and the joint stereo parameters, an approximation of the original selected channel B can be calculated.

25 Alternatively, the joint stereo module 140f can be implemented for performing binaural cue coding.

In the case of BCC, the joint stereo module 140f is operative to output the channel side information such that the 30 channel side information are quantized and encoded ICLD or ICTD parameters, wherein the selected original channel reserves as the actual to be processed channel, while the respective downmix channel used for calculating the side in-

formation, such as the first, the second or a combination of the first and second downmix channels is used as the reference channel in the sense of the BCC coding/decoding technique.

- 5 Referring to Fig. 4, a simple energy-directed implementation of element 140f is given. This device includes a frequency band selector 44 selecting a frequency band from channel A and a corresponding frequency band of channel B.
- 10 Then, in both frequency bands, an energy is calculated by means of an energy calculator 42 for each branch. The detailed implementation of the energy calculator 42 will depend on whether the output signal from block 40 is a sub-band signal or are frequency coefficients. In other implementations, where scale factors for scale factor bands are
- 15 calculated, one can already use scale factors of the first and second channel A, B as energy values  $E_A$  and  $E_B$  or at least as estimates of the energy. In a gain factor calculating device 44, a gain factor  $g_B$  for the selected frequency band is determined based on a certain rule such as
- 20 the gain determining rule illustrated in block 44 in Fig. 4. Here, the gain factor  $g_B$  can directly be used for weighting time domain samples or frequency coefficients such as will be described later in Fig. 5. To this end, the
- 25 gain factor  $g_B$ , which is valid for the selected frequency band is used as the channel side information for channel B as the selected original channel. This selected original channel B will not be transmitted to decoder but will be represented by the parametric channel side information as
- 30 calculated by the calculator 14 in Fig. 1.

It is to be noted here that it is not necessary to transmit gain values as channel side information. It is also suffi-

cient to transmit frequency dependent values related to the absolute energy of the selected original channel. Then, the decoder has to calculate the actual energy of the downmix channel and the gain factor based on the downmix channel energy and the transmitted energy for channel B.

Fig. 5 shows a possible implementation of a decoder set up in connection with a transform-based perceptual audio encoder. Compared to Fig. 2, the functionalities of the entropy decoder and inverse quantizer 50 (Fig. 5) will be included in block 24 of Fig. 2. The functionality of the frequency/time converting elements 52a, 52b (Fig. 5) will, however, be implemented in item 36 of Fig. 2. Element 50 in Fig. 5 receives an encoded version of the first or the second downmix signal  $Lc'$  or  $Rc'$ . At the output of element 50, an at least partly decoded version of the first and the second downmix channel is present which is subsequently called channel A. Channel A is input into a frequency band selector 54 for selecting a certain frequency band from channel A. This selected frequency band is weighted using a multiplier 56. The multiplier 56 receives, for multiplying, a certain gain factor  $g_b$ , which is assigned to the selected frequency band selected by the frequency band selector 54, which corresponds to the frequency band selector 40 in Fig. 4 at the encoder side. At the input of the frequency time converter 52a, there exists, together with other bands, a frequency domain representation of channel A. At the output of multiplier 56 and, in particular, at the input of frequency/time conversion means 52b there will be a reconstructed frequency domain representation of channel B. Therefore, at the output of element 52a, there will be a time domain representation for channel A, while, at the

output of element 52b, there will be a time domain representation of reconstructed channel B.

It is to be noted here that, depending on the certain implementation, the decoded downmix channel Lc or Rc is not played back in a multi-channel enhanced decoder. In such a multi-channel enhanced decoder, the decoded downmix channels are only used for reconstructing the original channels. The decoded downmix channels are only replayed in lower scale stereo-only decoders.

To this end, reference is made to Fig. 9, which shows the preferred implementation of the present invention in a surround/mp3 environment. An mp3 enhanced surround bitstream is input into a standard mp3 decoder 24, which outputs decoded versions of the original downmix channels. These downmix channels can then be directly replayed by means of a low level decoder. Alternatively, these two channels are input into the advanced joint stereo decoding device 32 which also receives the multi-channel extension data, which are preferably input into the ancillary data field in a mp3 compliant bitstream.

Subsequently, reference is made to Fig. 7 showing the grouping of the selected original channel and the respective downmix channel or combined downmix channel. In this regard, the right column of the table in Fig. 7 corresponds to channel A in Fig. 3A, 3B, 4 and 5, while the column in the middle corresponds to channel B in these figures. In the left column in Fig. 7, the respective channel side information is explicitly stated. In accordance with the Fig. 7 table, the channel side information  $l_1$  for the original left channel L is calculated using the left downmix channel

Lc. The left surround channel side information  $ls_i$  is determined by means of the original selected left surround channel Ls and the left downmix channel Lc is the carrier. The right channel side information  $ri$  for the original right channel R are determined using the right downmix channel Rc. Additionally, the channel side information for the right surround channel Rs are determined using the right downmix channel Rc as the carrier. Finally, the channel side information  $ci$  for the center channel C are determined using the combined downmix channel, which is obtained by means of a combination of the first and the second downmix channel, which can be easily calculated in both an encoder and a decoder and which does not require any extra bits for transmission.

Naturally, one could also calculate the channel side information for the left channel e. g. based on a combined downmix channel or even a downmix channel, which is obtained by a weighted addition of the first and second downmix channels such as 0.7 Lc and 0.3 Rc, as long as the weighting parameters are known to a decoder or transmitted accordingly. For most applications, however, it will be preferred to only derive channel side information for the center channel from the combined downmix channel, i.e., from a combination of the first and second downmix channels.

To show the bit saving potential of the present invention, the following typical example is given. In case of a five channel audio signal, a normal encoder needs a bit rate of 30 64 kbit/s for each channel amounting to an overall bit rate of 320 kbit/s for the five channel signal. The left and right stereo signals require a bit rate of 128 kbit/s. Channels side information for one channel are between 1.5

and 2 kbit/s. Thus, even in a case, in which channel side information for each of the five channels are transmitted, this additional data add up to only 7.5 to 10 kbit/s. Thus, the inventive concept allows transmission of a five channel 5 audio signal using a bit rate of 138 kbit/s (compared to 320 (!) kbit/s) with good quality, since the decoder does not use the problematic dematrixing operation. Probably even more important is the fact that the inventive concept 10 is fully backward compatible, since each of the existing mp3 players is able to replay the first downmix channel and the second downmix channel to produce a conventional stereo output.

Depending on the application environment, the inventive 15 method for processing or inverse processing can be implemented in hardware or in software. The implementation can be a digital storage medium such as a disk or a CD having electronically readable control signals, which can cooperate with a programmable computer system such that the inventive 20 method for processing or inverse processing is carried out. Generally stated, the invention therefore, also relates to a computer program product having a program code stored on a machine-readable carrier, the program code being adapted for performing the inventive method, when the 25 computer program product runs on a computer. In other words, the invention, therefore, also relates to a computer program having a program code for performing the method, when the computer program runs on a computer.